

## BIROn - Birkbeck Institutional Research Online

Dewaele, Jean-Marc and Mercer, S. and Talbot, K. and von Blanckenburg, M. (2020) Are EFL pre-service teachers' judgment of teaching competence swayed by the belief that the EFL teacher is a L1 or LX user of English? *European Journal of Applied Linguistics* 9 (2), pp. 259-282. ISSN 2192-953X.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/31075/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

# **Are EFL pre-service teachers' judgment of teaching competence swayed by the belief that the EFL teacher is a L1 or LX user of English?<sup>1</sup>**

Jean-Marc Dewaele, Birkbeck, University of London

Sarah Mercer, University of Graz

Kyle Talbot, University of Graz

Max von Blanckenburg, Ludwig-Maximilians Universität München

## **1 Introduction**

The problematic issues surrounding Native-speakerism and the use of the terms native/non-native speaker (NS/NNS) have increasingly been the focus of academic work and debate (Aneja 2016; Holliday 2015; Kamhi-Stein 2016; Leonard 2018; Mahboob 2018; Richard 2017; Swan et al. 2015). Though the concept and the terms have been heavily criticized because they are simplistic and static (Faez 2011; Leonard 2018), researchers face a dilemma “because there is a necessity to use terms, ‘non-native speaker’ and ‘native speaker’, which should not be in use at all” (Holliday 2015: 12). Paradoxically, use of the terms even when defending the rights of the NNS may have contributed to perpetuating the use of the term. Brain and Selvi (2018: 1) explain that the movement of ‘non-native English-speaking teachers’, known as NNESTs, “emerged to counter the discriminatory practices in TESOL, and to establish a professional milieu characterized by such values as democracy, justice, equity, participation, and professionalism”. In his reflection on the implications and directions of NNEST research and action, Mahboob (2018: 1) presents it as a lens “through which NNESTs—as classroom practitioners, researchers, and teacher educators—take diversity as a starting point, rather than as a result”. It aims to break “the monolingual and/or native speaker biases in the field” (Mahboob 2018: 1).

While we fully agree with the aims of NNEST research, we have reservations about the label “NNEST”. Using a word that has accumulated increasingly strong negative connotations

---

<sup>1</sup> Pre-print of Dewaele, J.-M., Mercer, S., Talbot, K.R. & von Blanckenburg, M. (2020) Are EFL pre-service teachers' judgment of teaching competence swayed by the belief that the EFL teacher is a L1 or LX user of English? *European Journal of Applied Linguistics* Doi: 10.1515/eujal-2019-0030

inevitably has pragmatic and ideological consequences. To illustrate, consider the historical example of the gradual disappearance of the word “negro” which was widely accepted in the US until the late 1960s when it fell out of favour because of its association with the history of slavery and acts of discrimination. It was replaced by more neutral terms like “African American” and “Black” (Smith 1992) and the word “negro” became increasingly offensive to many in the US, even if no offense was intended. The terms ‘NNS’ and ‘NS’ have a comparable future ahead. There have been many calls over the years to abandon these terms (Paikeday 1985; Rampton 1990) but they have remained in use because there were no viable alternatives (Holliday 2015). However, an alternative set of terms has been proposed, namely, first (L1) and foreign language (LX) users, which is free of ideological connotations of superiority or prestige of one group over the other (Dewaele 2018) and offers a fresh and neutral perspective to investigate differences.

In this study, we are specifically interested in exploring to what extent the terms ‘NS/NNS’ may elicit underlying biases among pre-service teachers in Germany and Austria judging a video-recording of an EFL teacher after being told that the teacher was a “native speaker” or a “non-native speaker”<sup>i</sup>. Would the labelling of the teacher they were watching influence their judgment?

## **2 Literature review**

Davies (2003) wondered why the term ‘NS’ remained widely used in EFL circles, despite the impossibility of defining the concept and deciding reliably on who belongs, or who does not belong to that category. The term ‘NS’ is discriminatory, as it can be used to deliberately exclude speakers of certain varieties of a language or speakers who were not born with the language. Considering ESL/EFL teaching in the United States, Mahmood (2004) questioned how the cherished American values of equality could be violated in the field EFL/ESL by treating NNS teachers of English as unwanted. He identified programme administrators as part of the problem, who often harboured the erroneous belief that students prefer NS teachers. This belief has been branded “Native-speakerism” to describe an ideology that has the NS fallacy at its core. NS are perceived to be “the best models and teachers of English because they represent a ‘Western culture’ from which spring the ideals both of English and of the methodology for teaching it” (Holliday 2005: 6). Those who use the terms NS and NNS argue that these are neutral; however, this denies their

ideological underpinnings. For Holliday (2015: 11), Native-speakerism is a “wide-spread cultural disbelief – a disbelief in the cultural contribution of teachers who have been labelled ‘non-native speakers’”.

This disbelief can have a negative impact on how ‘NNS’ teachers are perceived by colleagues and students leading to discriminatory employment practices and lower pay. Holliday (2015) points out that Native-speakerism also affects ‘NS’ teachers who feel commodified, i.e., reduced to a list of commercial attributes, which typically do not include their professional training and experience. ‘NS’ teachers are frequently portrayed as global backpackers with little professional didactic or pedagogical competence beyond being born speaking a specific language and ‘NNS’ teachers are considered to be more competent in grammar but lacking in idiomatic and cultural knowledge (Freeman 2016; Medgyes 1994; Talbot and Mercer 2018). Both stereotypical views diminish the professionalism and competences that many teachers work hard to achieve whatever their language profiles and can indeed be damaging in the workplace in very practical ways (Richards 2017). Furthermore, there have been several discussions of teachers and their teaching styles which claim different approaches to didactics and pedagogy depending on whether the person is an L1 user of English or not (see, e.g., Medgyes 1994; Richards 2017; Shin 2008; Walkinshaw and Oanh 2014). Labelling teachers and their competences according to their language heritage can further perpetuate myths about language teacher competences which can have discriminatory consequences (Brain and Selvi 2018; Clark and Paran 2007; Llurda 2009; Medgyes 1994). It reflects the classic problem in research of essentializing people to a label and category that may have little bearing on the variables being measured but to which any perceived difference is then ascribed.

Medgyes (1994) noted that the use of the terms ‘NS/NNS’ was wide-spread. Indeed, in a study carried out within the UK, 72.3% of respondents still claimed that the “native English speaker criterion” was moderately or very important in hiring practices (Clark and Paran 2007). In a recent CLIL-related study in Austria, parents were seen to exert a considerable pressure on the institutions to employ “native speaker teachers” suggesting that the problems with these terms and associated prejudices are not necessarily inherent in the schools but stem from pressures in society more widely (Gruber et al., in press). In the same study, the perceived pressures of the label “NNS” and the perceived untenable comparison to supposed “NS” even led to some teachers quitting their CLIL teaching role (Gruber et al., in press). Other studies have also shown

how these labels can be potentially damaging to teacher self-efficacy and can lead to problems in acquiring or maintaining employment as educators (Clark and Paran 2007; Shin 2008). As Davis (2004: 440) explains, a person's sense of self can be threatened by "not being valued for one's self (one's language is perceived as not good enough), of someone else's language being presented not just as different (so much is obvious), but as better than yours, and of the pervading feeling that whatever you do you will never achieve "proper" command over the incoming language".

Dewaele (2018) has argued that the continued use of the terms 'NS/NNS' perpetuates the fallacy. Instead, he proposed a more neutral dichotomy 'L1 user' versus 'LX user', which is more inclusive as it includes people who might read or understand a language without actually speaking it. It also avoids any mention of level of proficiency and removes the deficit connotations for the LX user and is thus less likely to risk being misleading, inaccurate or discriminatory. This terminology also allows a more holistic perspective on language users, since multilinguals are L1 users as well as LX users.

The present study aims to examine whether the labels 'NS/NNS' triggered any stereotypes and bias associated with the terms among pre-service EFL teachers. In social stereotyping, people have expectations, prejudices and stereotypes associated with a social group based on a category. Based on this category, people generalize and make judgments and predictions about members of that social group. Such stereotyping may be conscious (explicit) but also unconscious (implicit). Implicit bias refers to the beliefs we may hold but be unaware of which influence our behaviours and judgments of others. In order to measure such implicit attitudes, it is important to use research methodologies that avoid direct self-report data, given how susceptible such explicit responses are to social desirability influences. Instead, methods are needed which take a more indirect approach to revealing underlying attitudes. This is particularly important when attitudes examined may reflect certain socially desirable or sensitive perspectives. For example, several implicit bias studies have examined whether discriminatory beliefs influence employment practices based on race and ethnicity even without the employee being aware of making such judgments (see, e.g., Rooth 2010; Ziegert and Hanges 2005); the parallels for potentially prejudiced practices within language education based on the category of NS/NNS are immediately apparent.

In this study, we wanted to investigate possible implicit bias among pre-service teachers and whether it may be affecting their judgments about teacher competence. We were cautious about whether the participants might already be conscious of the issues surrounding the discussion of Native-speakerism, even though we deliberately chose students at the start of their studies who had not yet been exposed to this topic within their studies. The risk is that students could be susceptible to social desirability bias and thus would provide the response they feel they ‘ought’ to give in a regular self-report survey. This meant we needed a methodological design which would reveal the kinds of deeply-held beliefs and implicit bias we wished to check for. Inspired by studies such as Darley and Gross (1983), we sought to investigate whether the labels ‘NS/NNS’ were linked to any underlying prejudices and stereotypes which could affect how our participants rated and evaluated the performance of an EFL teacher depending on the respective label used. The teacher being evaluated in each case was the identical same person, but the only difference was the priming through a different label (NS or NNS) used to describe the teacher.

### **3 Research questions**

- 1) Do participants rate the teacher differently on the dimensions of Language, Teaching, Assessment, Communication and a holistic judgment of Love of the Teacher when told she was an English NS rather than a NNS of English?<sup>ii</sup>
- 2) Does participants’ place of study and gender affect their teacher ratings?
- 3) What dimensions (Language, Teaching, Assessment, Communication) are the best predictors of the holistic judgment of Love of the teacher?

## **4 Methodology**

### **4.1 Participants**

A total of 266 trainee teachers (181 females, 80 males, 5 did not say) participated in the study. Their age ranged between 18 and 42 (*Mean* = 21.8 years, *SD* = 3.1). Participants were mostly Austrian (*n* = 133) and German (*n* = 113). A large majority thus had German as a first language (*n* = 254), sometimes combined with English (*n* = 8) or other languages. English was the most frequent second language (*n* = 225). The remaining participants had acquired English as a third or fourth language. The sample consisted of 53 bilinguals, 143 trilinguals, 57 quadrilinguals, 11

pentalinguals and 5 sextalinguals. In both Austria and Germany, the level of English of students starting their studies after school is at least B2 but many in this course would have been also C1. As such, it would have been no problem for them to express themselves in English in the questionnaire. Participants would also be familiar with the competence areas featured in the questionnaire as these form a part of their teacher education programme from the beginning on.

Slightly over half of participants were enrolled in Graz ( $n = 145$ ), with the other participants studying in Munich ( $n = 121$ ). In both countries, students study two subjects to become teachers in state schools. In addition to English, they were studying a wide range of subjects, including history ( $n = 43$ ), geography ( $n = 24$ ), German ( $n = 20$ ), biology ( $n = 16$ ) and 24 other subjects. A majority of participants hoped to become a teacher in secondary schools ( $n = 228$ ), with a smaller number aiming for primary ( $n = 26$ ) and tertiary education ( $n = 11$ ). Whether the participants responded to a NS or NNS prime depended on the class they were attending and each class was treated as an intact unit.

#### **4.2 Procedure stage 1**

The study was conducted in two parts. The first phase was to create a video for the participants to watch in the main part of the study. Four ELT teachers working in secondary schools in Austria volunteered to take part in stage 1. These teachers were known to the authors and approached specifically due to their high level of bilingualism in German and English. They were informed of the two stages of the study and that they may or may not be needed for the second stage. All were L1 German speakers. The aim was to find a teacher whose English was sufficiently ambiguous to be able to be identified as an L1 speaker of either English or German. These teachers recorded a short monologue which was listened to and rated anonymously by all four members of the research team and two external raters for whether they felt the teacher had L1 English or German or unable to say. The teacher whose English was rated as the most ambiguous was approached and agreed to have her teaching filmed for the second part of the study. A regular ELT class in an Austrian secondary school was filmed with the recording focusing on her person. None of the students were visible for ethical reasons. This video was shortened to an approximately 5-minute segment which showed the teacher in a variety of teaching situations during the one class. The teacher watched and agreed to the use of the video for research purposes. The video was not made otherwise available.

### 4.3 Procedure stage 2

University students at the beginning of their studies to become English teachers in Austria and Germany were asked to take part in the study during their introductory courses to language teaching. They were informed that participation was voluntary, and they could choose not to complete the questionnaire if they wished. Students were merely informed that they would be asked to evaluate their first impressions of a teacher's competences based on a video they would watch. They were not yet informed of the underlying rationale behind the study, although this was explicitly made clear in a deliberate debrief after all the data were collected. The key design feature was that approximately half of the students (in intact groups) were told orally and at the top of the questionnaire that they would watch and rate a "native-speaker teacher" ( $n = 131$ ) for the first impressions they would gain of her and her teaching from watching a brief segment. The other students ( $n = 135$ ) were told they were watching a "non-native speaker teacher". Naturally, both groups were watching the same identical video. The only difference was whether it was framed as being of a "native speaker teacher" or "non-native speaker teacher". This priming was done with the intention of examining whether this difference alone would influence the evaluations of the teacher and thereby reveal any possible underlying implicit bias associated with the terms.

In the instructions to students, it was stressed that this rating of her teaching would be impressionistic in part and we were not looking for an accurate rating, especially as some elements were included for evaluation that were infrequent in the short segment. The focus was on the overall impression they formed of her as a teacher based on the short film extract of her actual teaching. For a number of students, this instruction was problematic although certainly not for all. In the open-ended data, 17 students commented explicitly that it was either difficult to rate things they had not seen or otherwise challenging to form judgments about the teacher based on a first impression alone. The video was deliberately only shown once as implicit attitudes are revealed more on instinctive, spontaneous responses than through careful, extended reflection.

Finally, it is important to note for ethical purposes that after data had been collected, all the teachers were asked to debrief their students about the underlying purposes of the study and issues surrounding the use of these terms. We wanted to ensure that the study did not inadvertently further propagate or strengthen the damaging attitudes inherent in the terms 'NS/NNS'. One researcher who gathered data for this study from her class reported that this teaching session



created an extremely interesting and learning-rich opportunity. Learners were encouraged to ask questions and discuss the aims and hypotheses of this study. It led to thought-provoking and valuable discussions about issues surrounding these terms and it is hoped that they will learn from the experience as they move forward in their careers. For those wishing to replicate this study, we consider this a vital stage to include for ethical reasons.

#### 4.4 Instrument

After providing the background information reported above, participants watched an authentic video of an English teacher teaching her students in a regular secondary school in Austria. The questionnaire was entitled *Evaluating language teacher competences* and each group was also primed orally about whether they would be watching a “non/native speaker” English teacher.

The written instructions on the questionnaire were as follows:

You are now going to watch a five-minute clip showing a **non/native speaker** of English [respective group] teaching students in a secondary school in Austria. Based on the first impression you gain from watching her teach for just these few minutes, we would like you to evaluate her teaching competences in five key categories. There are no right or wrong answers. We are interested in the impressions you gained from this brief extract from a lesson and would like you to answer based on what you believe is most likely true for this teacher. Please note that your answers will not have any effect on or consequences for the teacher who is happy to have her teaching looked at by you. Many thanks for your input.

Participants were then presented with five sections containing a total of 27 statements with 6-point Likert scales with the following choices: Strongly Disagree (1), Disagree (2), Slightly Disagree (3), Slightly Agree (4), Agree (5) and Strongly Agree (6). These statements and categories were based on a framework of language teacher competences developed by Eaquals, an international association for language education that advises the Council of Europe on language policy issues. The framework provides a comprehensive overview of key professional competences with detailed descriptors, formulated to support language professionals in self-assessment and evaluation (Eaquals 2016), and, as such, presented a suitable foundation for developing the questionnaire used in this study. The statements and categories were adapted to suit the context, purpose and specific video segment the students would watch but with the aim of being as comprehensive as possible and reflecting diverse practical classroom-based practices of language teachers. We were keen to elicit a nuanced evaluation of competences in different competence

areas and not only a cumulative evaluation of competences which could mask possible domain differences.

The first section was entitled **Language** which contained 8 items. It started with: “The teacher can respond to all issues concerning communicative speaking instruction for learners up to C2 level”. The following items substituted “speaking” with listening, writing, reading, pronunciation, grammar, vocabulary. The section concluded with the item: “The teacher can explain distinctions and the meaning of different collocations, idiomatic expressions and connotations of lexical items”. Cronbach alpha values are presented in table 1.

The second section was entitled **Teaching** and contained the following 10 items: 1) The teacher can decide on appropriate sequences of activities, timing and pace for lesson phases and activities. 2) The teacher can match learning activities to learners’ language level and needs. 3) The teacher can use various resources effectively, including the board, technology, voice, and body language. 4) The teacher can effectively and flexibly use a broad range of teaching approaches and techniques. 5) The teacher can efficiently and flexibly set up and monitor whole class and pair, group, and individual work. 6) The teacher can give clear, comprehensible instructions. 7) The teacher can establish rapport and positive group dynamics with learners. 8) The teacher can handle groups of learners from different cultural backgrounds, of different ages, and/ or with different abilities and needs. 9) The teacher can create an atmosphere of respect, tolerance and understanding. 10) The teacher can motivate learners effectively.

The third section entitled **Assessment** contained three items: 1) The teacher can provide individualised and relevant feedback to enable the learners to monitor their own progress and achievement. 2) The teacher can recognise and diagnose diverse learners’ errors. 3) The teacher can use a range of techniques for responding to and offering advice on learners’ errors.

The fourth section **Communication** contained 5 items: 1) The teacher can communicate effectively with learners of the target language at all levels. 2) The teacher can identify and overcome communication difficulties and maintain communication flow. 3) The teacher can ensure that learners understand the relevance of cultural conventions such as proximity, politeness, punctuality, directness etc. 4) The teacher can systematically develop learners’ ability to reflect on cultural similarities and differences. 5) The teacher can help learners to reflect on cultural behaviour, traditions, artefacts etc. with materials and activities appropriate to the group.

A single closed item eliciting a holistic judgment concluded the questionnaire: “I would love to have this person as my teacher”. It was followed by an open question: Why? / Why not? Responses generated a corpus of 3888 words from 237 respondents. These data were first open content coded by one of the authors. A second author then applied a second wave of more refined thematic content coding of the data. In this wave of coding, each of the 237 responses was first coded as either from a questionnaire about NS or NNS. Subsequent codes were layered on top of the NS or NNS codes. This allowed us to observe co-occurring codes in the data and whether certain topics were addressed more in responses regarding supposed NS or NNS teachers (e.g., the code *humour* was coded 15 times for NNS and 12 times for NS). These qualitative data were analysed for salient themes, particular responses in the questionnaires regarding NS/NNS, teacher characteristics as well as the research method itself.

#### **4.5 Data**

A one-sample Kolmogorov-Smirnov test revealed that the distribution of scores on the four dimensions and the item “Love of the teacher” was skewed toward the higher end of the scale, thus not being entirely normally distributed (see table 1). However, calculation of Q-Q plots (figures 1, 2, 3, 4 and 5) suggests that the scores follow a normal distribution reasonably well except for the extreme tail of Language and Teaching (values below 4) and for the extreme tail of Assessment, Communication and Love of the teacher (values below 3). We thus opted for the more powerful parametric statistics. A Cronbach alpha analysis revealed that the four dimensions have sufficient internal consistency. Descriptive statistics are presented in table 1.

**Table 1.** Descriptive statistics

Variable	Min	Max	Mean	SD	Kolmogorov	Cronbach al-
					Smirnov	pha
Language (8 items)	3.1	6	5.03	0.61	0.860**	0.861
Teaching (10 items)	3.1	5.9	4.97	0.59	0.112**	0.844
Assessment (3 items)	1	6	4.63	0.80	0.131**	0.746
Communication (5 items)	3	6	4.80	0.66	0.096**	0.801
Love Teacher (1 item)	1	6	4.90	1.03		

\*\* $p < .01$

Figure 1: Q-Q plot of Language

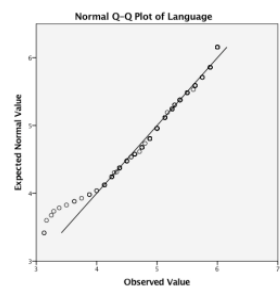


Figure 2: Q-Q plot of Teaching

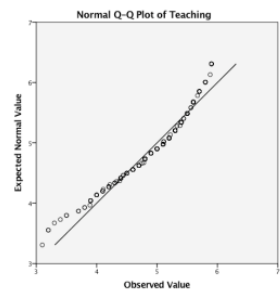


Figure 3: Q-Q plot of Assessment

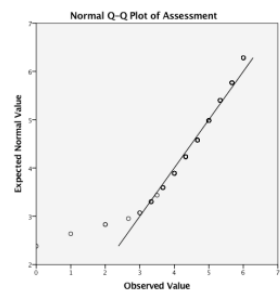


Figure 4: Q-Q plot of Communication

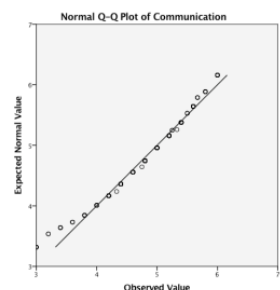
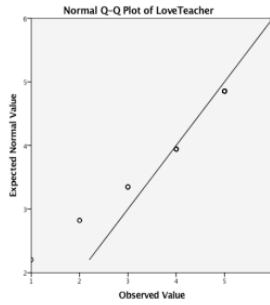


Figure 5: Q-Q plot of Love teacher



## 5 Results

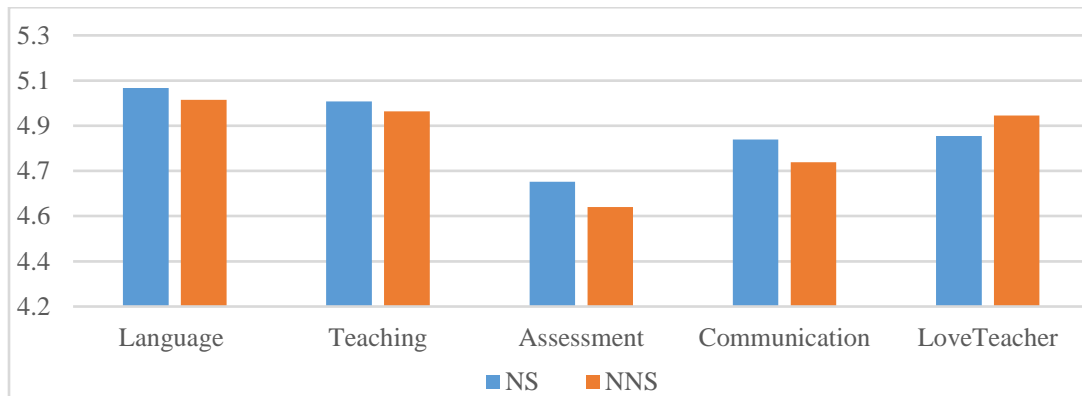
### 5.1 Quantitative analysis

An independent t-test revealed that being told that the teacher was a NS or an NNS of English had no effect at all on participants' teacher ratings on the four dimensions and their desire to have the person as a teacher (love teacher) (see table 2 and figure 6).

**Table 2.** The effect of believing that the teacher was a NS or NNS of English

Variable	<i>t</i>	<i>df</i>	<i>p</i>
Language	0.605	261	0.546
Teaching	0.528	261	0.598
Assessment	0.991	261	0.323
Communication	1.000	261	0.276
Love Teacher	0.621	257	0.535

Figure 6: The effect of believing that the teacher was a NS or NNS of English

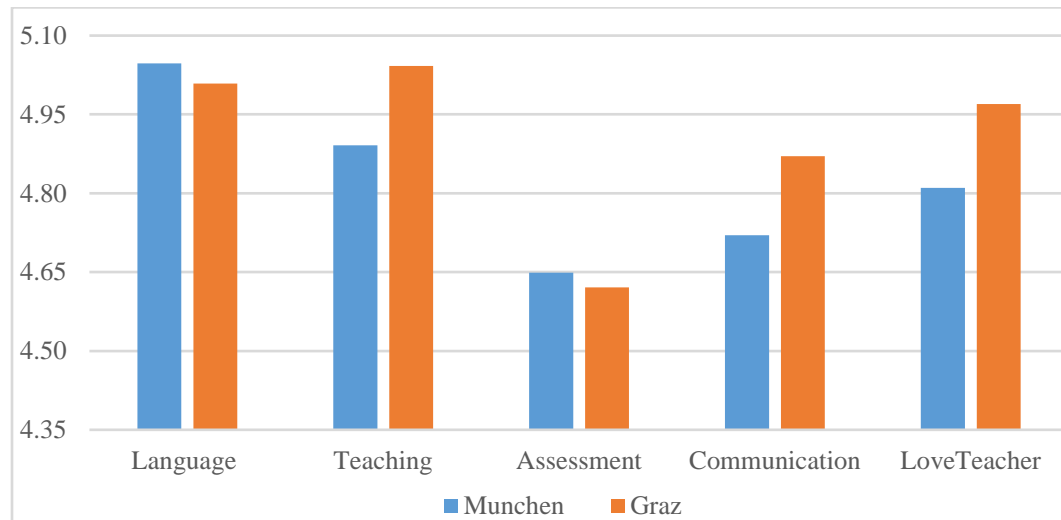


An independent t-test revealed that participants from Munchen rated the teacher significantly lower than their peers from Graz on Teaching (Cohen's  $d = 0.245$ , a small effect size according to Plonsky and Oswald 2014), and they rated the teacher marginally lower on Communication (see table 3).

**Table 3.** The effect of participants' place of study

Variable	<i>t</i>	<i>df</i>	<i>p</i>
Language	.509	264	.611
Teaching	-2.091	264	.038
Assessment	.283	264	.777
Communication	-1.873	264	.062
Love Teacher	-1.31	260	.190

Figure 7: The effect of participants' place of study



Participants' gender turned out to be completely unrelated to their judgment of the teacher's skills and their love for this teacher (see table 4).



**Table 4.** The effect of participants' gender

Variable	<i>t</i>	<i>df</i>	<i>p</i>
Language	1.000	259	0.261
Teaching	2.021	259	0.118
Assessment	0.933	259	0.352
Communication	0.833	259	0.405
Love Teacher	0.351	255	0.723

To identify the significant relationships between teacher dimensions and Love of the teacher, we ran preliminary Pearson correlation analyses (see table 5). All 4 dimensions were linked significantly with Love of the teacher, so all were included in a linear multiple regression analysis in order to find the strongest predictors. The regression analysis can remove redundancy from the predictor variables to see which are retained compared to the zero-order correlations.

**Table 5.** Pearson correlation analyses between the teacher dimensions and Love of the teacher (*N* = 262)

Variable	Love of teacher
Language	.500***
Teaching	.629***
Assessment	.391***
Communication	.500***

\*\*\*  $p < .0001$

A Pearson correlation analysis was run to check the degree of inter-correlation between the four independent variables (see table 6). The results show highly significant positive relationships with variables sharing between 23% and 46% of variance. Unsurprisingly, the highest correlation occurs between Teaching and Communication, as clear communication is a requisite to good teaching (see table 3).

**Table 6.** Inter-correlation between the independent variables

Variable	Language	Teaching	Assessment
Teaching	.531***		
Assessment	.489***	.494***	
Communication	.552***	.681***	.539***

\*\*\*  $p < .0001$

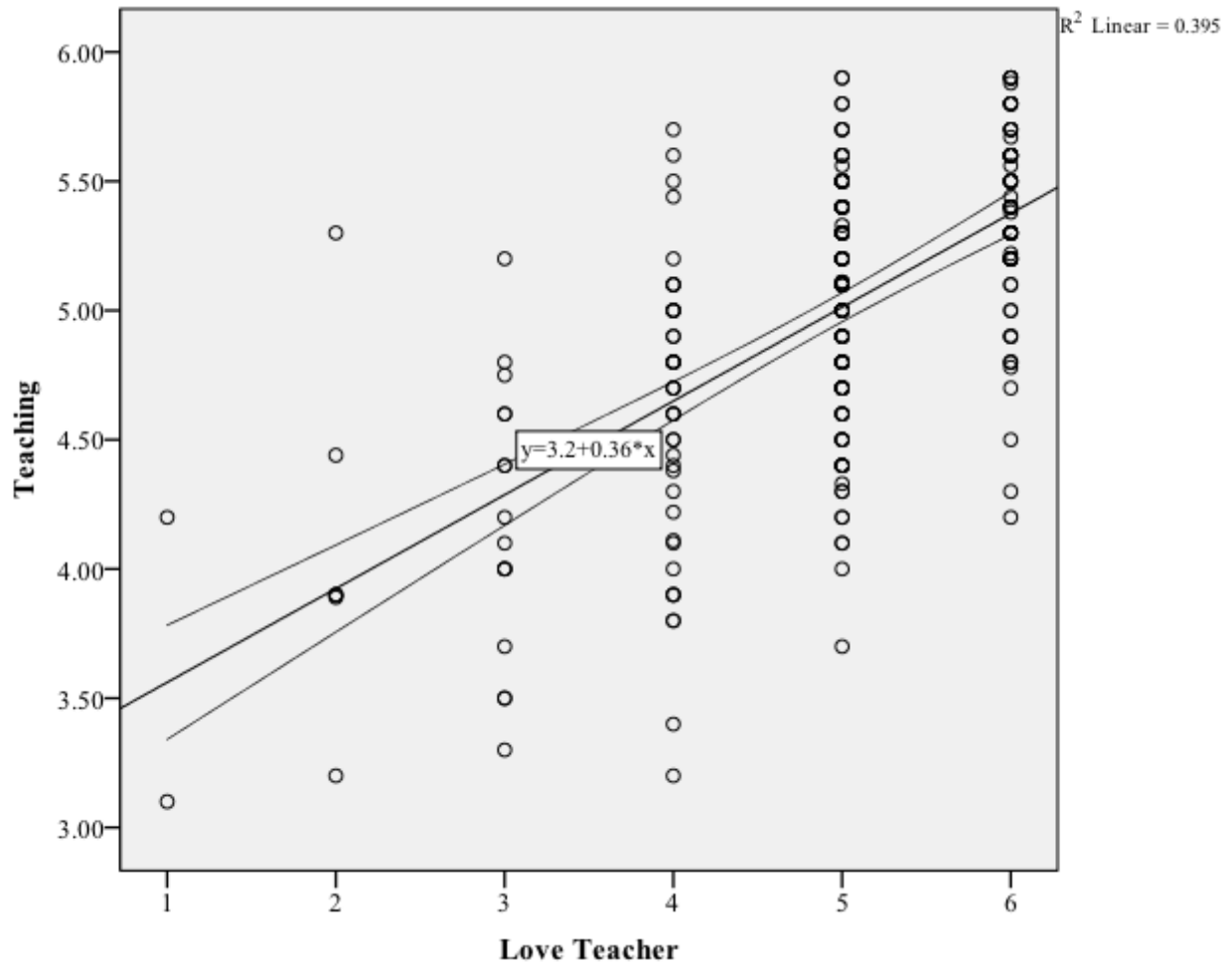
Green (1991) recommends a minimum sample size of 50 for any regression, with an additional 8 observations per term. This means the minimum sample size for 4 independent variables is 82, which is well below our sample size of 266.

Multiple regression analysis (enter method) was used. Values for the variance inflation factor (VIF), which quantifies the severity of multicollinearity, hover between 1.5 and 2.2, which is well below the recommended cut-off point of 5 (Kutner, Nachtsheim and Neter 2004). The Durbin Watson test for autocorrelation in the residuals was 1.953, which suggests there is no autocorrelation in the sample.

A significant regression equation was found for Love of teacher, indicating that two variables predicted 44% of the variance (Adjusted  $R^2 = .427$ ,  $F(4\ 257) = 49.6$ ,  $p < .0001$ ). The strongest predictors were Teaching ( $Beta = .470$ ,  $t = 7.05$ ,  $p < .0001$ ), followed by Language ( $Beta = .207$ ,  $t = 3.45$ ,  $p < .001$ ). Assessment and Communication did not predict any unique variance in Love of teacher. In other words, participants' love of the teacher was mainly determined by their judgment about the teacher's teaching skills, with her language skills as a much weaker predictor. The overall effect size could be described as medium-to-large (Plonsky and Ghanbar 2018).

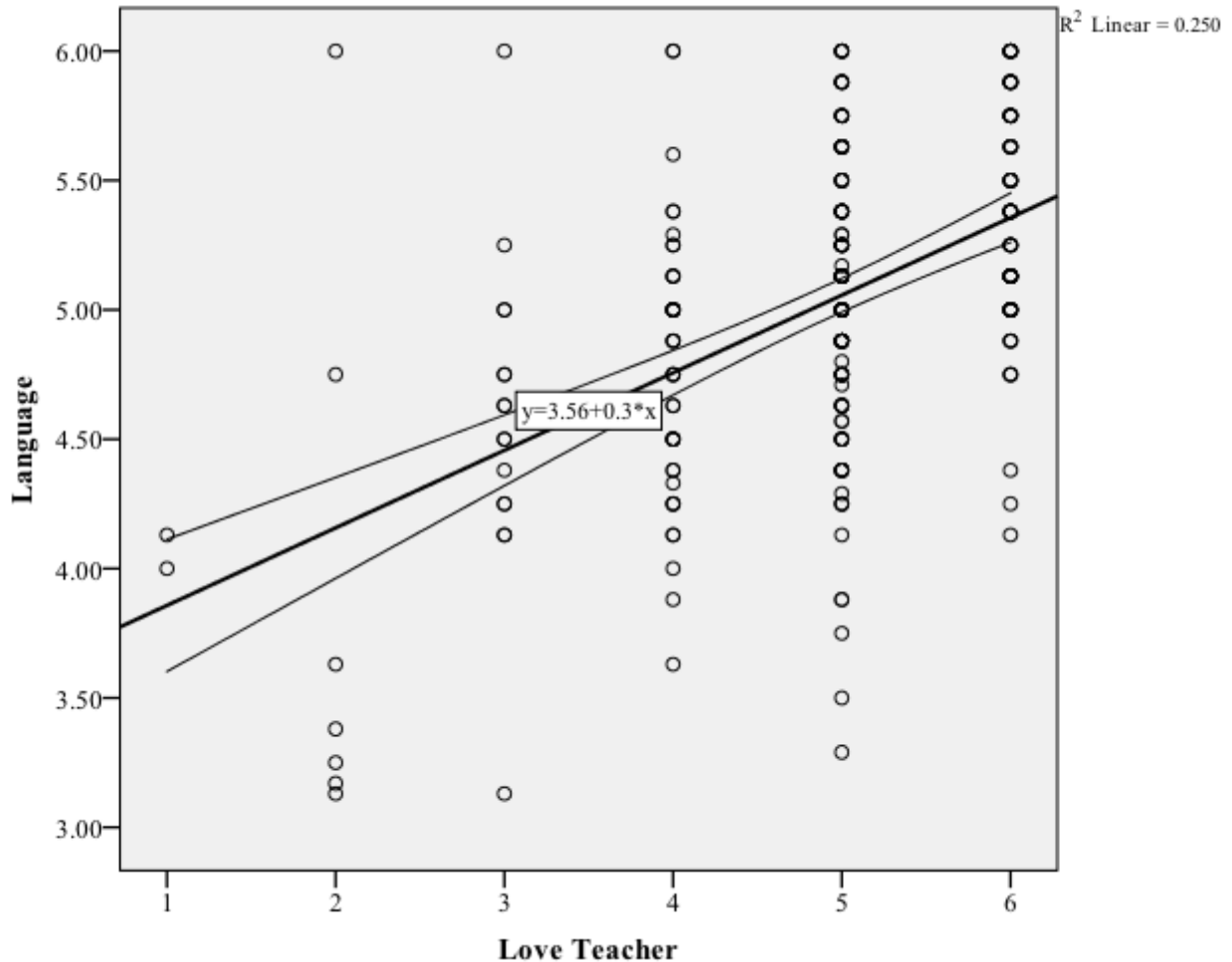
A first scatterplot with a regression line shows that higher Teaching scores are linked to increased values for Love of teacher (see figure 8).

Figure 8: The relationship between Teaching scores and Love of teacher (with 95% Confidence interval)



A similar pattern emerges in the second scatterplot with a regression line shows that higher Language scores are linked to higher values for Love of teacher (see figure 9).

Figure 9: The relationship between Language scores and Love of teacher (with 95% Confidence interval)



## 5.2 Qualitative analysis

The analysis of the qualitative data revealed some interesting nuances. A small number of respondents (14 out of 266) mentioned explicitly that the teacher was a NS or NNS, which suggests that the priming was noticed at least by some students. Six commented merely that it was a NS as explanation for why they liked the teacher. Five commented explicitly on some aspect of her language competence positively because she was a NS. For example, one respondent wrote, “It's always a positive thing to have someone being raised with a certain language teaching it. That person would know which phrases are used more commonly and how slang conversations

work”. However, one respondent felt the teacher’s ‘nativeness’ may be a problem for young learners at lower levels. One person reported not liking the teacher’s pronunciation (“She sounds very German-English”), though she also said this teacher explained content well and seemed competent. Two individuals claimed that having a NS teacher was more motivating or engaging for students: “Having a native speaker teaching is generally a nice thing for me, as students instantly show more motivation/ interest”. At least for some of these learners, there still were certain expectations of competence related to a supposed ‘NS teacher identity’ and/or beliefs about supposed benefits of having an L1 English teacher. As one respondent explained, “The fact that she’s a native speaker is great. As a student at school I’d feel like she’s very authentic and capable”. With the exception of the student who commented explicitly about not liking this teacher’s pronunciation, there were no instances in the qualitative data of any disparaging or otherwise negative comments specifically about her as a supposed NNS teacher.

When examining the descriptions of the teacher, we also looked at whether there were any differences according to whether these descriptions were located in a NS or NNS questionnaire, even if the responses did not explicitly cite this as a reason. Here, one notable finding was that all 7 instances of describing the teacher as ‘authentic’ occurred in data sets from the ‘NS’ primed group. The description of competence or effectiveness as a teacher was virtually the same across both groups ( $n = 12$  in the NNS condition;  $n = 14$  in the NS condition). Interestingly, an evaluation of her positivity in her approach to teaching was more marked among the NNS responses ( $n = 23$  NNS;  $n = 10$  NS).

On the whole, the comments were largely positive about the teacher. There were only 51 instances of critical feedback overall and these were mostly centred on classroom management and the perception of the teacher as being hurried or stressed and these were evenly distributed among the NS/NNS questionnaires ( $n = 26$  NNS,  $n = 25$  NS). The remaining comments all focused on the positive characteristics of this teacher. These included her pedagogy, choice of classroom topics, positivity, personality, pedagogical caring, motivational role, ability to engage students, classroom competence and effectiveness as a teacher, among other factors.

## 6. Discussion

The answer to RQ1 is satisfying, in that participants' opinion of the teacher they were watching in action were not swayed by believing that she was a NS/L1 or NNS/LX user of English. In other words, they seem not to have been afflicted by Native-speakerism as we had feared (Holliday 2015; Holliday and Aboshiha 2009) and they generally did not perceive the teachers with different labels as two different types of teacher (cf. Aslan and Thompson 2016). However, the open-ended data does suggest that some individuals did still hold certain prejudices about teachers depending on whether they believe them to be NS/L1 or NNS/LX users of English, although these were notably in very low numbers. Our hope is that educating the next generation of teachers consciously and explicitly about this issue and the related problems will inform attitudes from a grassroots level. It is therefore not impossible that the aims of the NNEST movement may be achieved, in Europe, within the next generation of teachers (cf. Brain and Selvi 2018; Leonard 2018; Mahmood 2018).

Naturally, as the participants in this study are training to be EFL teachers, it might be expected that they would have a more open opinion and less prejudice than others, given their own status and identities. In the debrief sessions, the students seemed genuinely surprised to hear of Native-speakerism, although we know from other research that the terms and associated problems do still exist in educational contexts (Aneja 2016; Holliday 2015; Kamhi-Stein 2016, Shin 2008; Richards 2017). However, the positive findings in these data do raise questions about the extent to which the values and stereotypes typically associated with the terms are present in certain populations, more so than others. From our experience in other studies (e.g., Gruber et al., in press), we know that parents are putting a notable pressure on schools to employ so-called 'NS teachers'. It would be of interest to replicate this study with those less directly involved in language or education in anyway such as parents with other professions and backgrounds, and see whether the findings might have been different. In addition, it would be valuable to conduct this study in different settings. Many countries have different histories and experiences with language education and the role of L1 users of English and, as such, it would be valuable to replicate this study in settings which differ notably to this setting such as, for example, Vietnam (Canh and Renandya 2017).

The answer to RQ2 is mainly negative as participants' place of study and gender had a limited or no effect. The Austrian and German education contexts are very similar, and this was one reason the study was conducted in both settings in the expectation that the attitudes and experiences would be comparable. However, in-group bias refers to humans' tendency to rate people similar on some characteristic of self as more highly than others from different social groups. In education, for example, there is some research with mixed results which suggests that students may rate other teachers more positively when they are of the same gender (Boring, Ottoboni and Stark 2016), although this was found not to be the case in this study. In future work, it would be worth examining also students' previous experiences of teachers and whether their responses may be affected by the perceived similarity of the teacher being evaluated to a previous English teacher they had experienced.

RQ3 focused on the identification of predictors of scores on the final holistic question whether participants would love to have the person they just watched as their teacher. The strongest predictor turned out to be Teaching, which is not surprising considering that the participants were studying to become teachers and the research was conducted in the context of a teacher education class. In other words, they focused specifically on the various teaching-specific skills deployed by the teacher in the video. This is further supported by the qualitative data. A majority of comments focused on the teacher's pedagogy, although a large number of comments also referred to her personality. Language came as a weaker second predictor. This suggests that participants judged that sufficient mastery of English was a crucial aspect of their willingness to choose a specific English teacher, but it was not the most important aspect. The finding fits with findings in Dewaele, Franco Magdalena and Saito (2019) on the effect of teacher characteristics on FL Enjoyment and Anxiety of 210 Spanish EFL learners that teacher's friendliness was the strongest predictor of FL Enjoyment, while a teacher's foreign accent was a weaker negative predictor. Assessment and Communication were not significant predictors in the present study though they were positively correlated with Love of the Teacher. It is likely that these two dimensions were absorbed into the more general Teaching dimension. In other words, participants' love of the teacher was mainly determined by their judgments of her ability to teach in methodological and pedagogically pleasing ways as reflected in how she teaches, the group dynamics she is able to engender and her teaching persona more generally. It is important to reflect that these characteristics may simply reflect the aspects, which are more salient for this specific teacher's

style and the specific video extract, as opposed to being the dimensions the teachers would generally comment more widely on. The findings are encouraging for all pre-service teachers as these are aspects of teaching that an educator can work on, develop and improve which is an optimistic view for language educators and their professional development.

There are a number of limitations to this study. Firstly, our participants belong to a homogenous age group, a defined professional group and they have very similar language profiles. It would be of interest to replicate the study with other populations, in particular with parents from the wider population as well as internationally diverse language education contexts and across different languages. Ideally, such research could also then be combined with a programme of education and awareness-raising among the populations who took part as was the case in this study. The whole process turned out to be a valuable didactic opportunity. The second issue concerns the subtlety of the design and priming procedures, which could possibly have been too subtle to trigger noticing and implicit bias, although a limited number of responses, which explicitly referred to the teacher's language heritage, imply it was indeed noticed at least by some. Yet, it is worth reflecting on whether possibly the design and presentation of the teacher and priming use of the terms was in fact possibly missed by some of the participants. From the qualitative data, there were a small number of participants who displayed the types of prejudices or preferences we had perhaps anticipated and feared suggesting that in individual cases the stereotypes and bias associated with the terms do persist for some individuals, even if at the group level, there was less evidence of this. The study could in the future be extended to explore in more depth through interviews the responses of those who display some kinds of prejudices and bias in their open-ended responses to better understand the reasons and tenacity of such beliefs.

## **7. Conclusion**

This study took a quasi-experimental design to investigate possible implicit bias in the use of the terms 'NS/NNS' when framing the evaluation of a teacher's performance among pre-service EFL teachers. It was pleasing to see that this did not appear to be the case for this population although the qualitative data suggests that some individuals still may hold some biased or prejudiced attitudes or beliefs towards NNS/LX teachers, or, alternatively, positive stereotypes about NS/L1 teachers that were masked in the group-level data. In both cases, the effects of any bias



associated with these terms can be detrimental. Educating pre-service language teachers explicitly about these issues and the negative effects of continuing to use a ‘NS/NNS’ dichotomy would be a natural starting point for engendering change. However, we also feel a broader agenda of research combined with awareness-raising could reveal any underlying prejudices and lingering stereotypes in diverse areas of society and sensitise people to the possible effects of using these terms. We would like to be optimistic and assume that these terms do in fact not hold any implicit bias for people such as in our study, but we need further research evidence from across populations, languages and settings to elucidate whether and to what extent judgments of teachers may be being influenced by their language status (L1 or LX) and/or the use of the labels ‘NS/NNS’.

In terms of the positive evaluations of the teacher, her ability to communicate effectively and teaching style were defining for why the students would have wanted such a teacher themselves. Both results are encouraging in their potential for development and also the sense that the next generation of German and Austrian EFL teachers may have moved or at least be moving beyond the restrictive and discriminatory dichotomies of the past. Nevertheless, while we are pleased at the findings from this study, we do not feel there is cause for complacency. We feel that it is imperative to finally bury the NS myth (cf. Davies 2003; Holliday 2015; Llurda 2009; Paikeday 1985) and to do so by getting rid of the labels NS/NNS/NNEST across the teaching profession. The use of more neutral terms, which do not aim at essentialising people to a single restrictive and static label, indicate a new era characterized by a more open climate in which language diversity and change is appreciated and valued. Our profession needs to define professional competence in terms of linguistic proficiency and didactic and pedagogical competences, irrespective of the person’s language heritage.

**Acknowledgements:** We would like to thank all our participants and the reviewers for their excellent comments.

## 8. References

- Aneja, Geeta A. 2016. (Non)native speakerhood: Rethinking (non)nativeness and teacher identity in TESOL teacher education. *TESOL Quarterly* 50. 572-596. doi: 10.1002/tesq.315
- Aslan, Erhan & Amy S. Thompson. 2016. Are they really “two different species”? Implicitly elicited student perceptions about NESTs and NNESTs. *TESOL Journal* 8. 277–294. doi:10.1002/tesj.268
- Boring, Anne, Kellie Ottoboni & Philip B. Stark. 2016. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Braine, Georges & Ali Fuad Selvi. 2018. NNEST Movement. *The TESOL Encyclopedia of English Language Teaching*, 1–6. doi:10.1002/9781118784235.eelt0039
- Canh, Le Van & Willy A. Renandya. 2017. Teachers’ English proficiency and classroom language use: A conversation analysis study. *RELC Journal* 48. 67–81. Doi: 10.1177/0033688217690935
- Clark, Elizabeth & Amos Paran. 2007. The employability of non-native-speaker teachers of EFL: A UK survey. *System* 35. 407-430. Doi: 10.1016/j.system.2007.05.002
- Darley, John M. & Paget H. Gross. 1983. A hypothesis-confirming bias in labelling effects. *Journal of Personality and Social Psychology* 44. 20-33.
- Davies, Alan. 2003. *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters.
- Dewaele, Jean-Marc. 2018. Why the dichotomy ‘L1 Versus LX User’ is better than ‘Native Versus Non-native Speaker’. *Applied Linguistics*, 39, 236-240. Doi:10.1093/applin/amw055
- Dewaele, Jean-Marc, Andrea Franco Magdalena & Kazuya Saito. 2019. The effect of perception of teacher characteristics on Spanish EFL Learners’ Anxiety and Enjoyment. *The Modern Language Journal* 103. 412-427. Doi: 10.1111/modl.12555
- Eaquals 2016. The Eaquals Framework for Language Teacher Training & Development. <https://www.eaquals.org/our-expertise/teacher-development/the-eaquals-framework-for-teacher-training-and-development/>
- Faez, Farahnaz. 2011. Are you a native speaker of English? Moving beyond a simplistic dichotomy. *Critical Inquiry in Language Studies* 8. 378–99. Doi: 10.1080/15427587.2011.615708

- Freeman, Donald. 2016. *Educating Second Language Teachers*. Oxford: Oxford University Press.
- Green, Samuel B. 1991. How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research* 26. 499–510.
- Holliday, Adrian R. & Pamela Aboshiha. 2009. The denial of ideology in perceptions of ‘Nonnative Speaker’ teachers. *TESOL Quarterly* 43, 669-689.
- Holliday, Adrian R. 2005. *The struggle to teach English as an international language*. Oxford: Oxford University Press.
- Holliday, Adrian R. 2015. Native-speakerism: Taking the concept forward and achieving cultural belief. In Adrian R. Holliday, Pamela Aboshiha & Anne Swan (eds.), *(En)Countering Native-speakerism. Global Perspectives*. Basingstoke: Palgrave MacMillan, 11-25.
- Kamhi-Stein, Lía D. 2016. The non-native English speaker teachers in TESOL movement. *ELT Journal* 70. 180–189.
- Kay, Aaron C., Martin V. Day, Mark P. Zanna, A. David Nussbaum. 2013. The insidious (and ironic) effects of positive stereotypes. *Journal of Experimental Social Psychology* 49. 287–291. Doi: 10.1016/j.jesp.2012.11.003
- Kutner, Michael H., Christopher J. Nachtsheim & John Neter. 2004. *Applied linear regression models* (4th ed.). Boston: McGraw-Hill Irwin.
- Leonard, Josie. (2018). Beyond “(non) native-speakerism”: Being or becoming a native-speaker teacher of English. *Applied Linguistics Review*, doi:10.1515/applirev-2017-0033
- Llurda, Enric. 2009. The decline and fall of the native speaker teacher. In Vivian J. Cook and Li Wei (eds.), *Language Teaching and Learning*, 37-53. Oxford: Continuum.
- Mahboob, Ahmar. 2004. Demystifying the Native Speaker in TESOL. *SPELT Journal* 19. 1-14.
- Mahboob, Ahmar. 2018. NNEST Lens: Implications and Directions. *The TESOL Encyclopedia of English Language Teaching*, 1–7. doi:10.1002/9781118784235.eelt0038
- Medgyes, Peter. 1994. *The non-native teachers*. London: MacMillan.
- Paikeday, Thomas M. 1985. *The native speaker is dead!* Toronto: Paikeday Press.
- Plonsky, Luke & Frederick L. Oswald. 2014. How big is ‘big’?: Interpreting effect sizes in L2 research. *Language Learning* 64. 878–912. Doi:10.1111/lang.12079
- Plonsky, Luke & Hessameddin Ghanbar. 2018. Multiple regression in L2 research: A methodological synthesis and guide to interpreting R<sup>2</sup> Values. *The Modern Language Journal*. DOI:

10.1111/modl.12509

- Rampton, Ben. 1990. Displacing the 'native speaker': expertise, affiliation, and inheritance. *ELT Journal* 44. 97–101.
- Richards, Jack C. 2017. Teaching English through English: Proficiency, pedagogy and performance. *RELC Journal* 48. 7–30. Doi:10.1177/0033688217690059
- Rooth, Dan-Olof. 2010. Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics* 17. 523–534. [doi.org/10.1016/j.labeco.2009.04.005](https://doi.org/10.1016/j.labeco.2009.04.005)
- Shin, Sarah J. 2008. Preparing non-native English speaking ESL teachers. *Teacher Development* 12. 57–65. Doi: 10.1080/13664530701827749
- Smith, Tom W. 1992. Changing racial labels: from 'Colored' to 'Negro' to 'Black' to 'African American'. *Public Opinion Quarterly* 56. 496–514. Doi: 10.1086/269339
- Swan, Anne, Pamela Aboshiha & Adrian Holliday. 2015. *(En)Countering Native-speakerism: Global Perspectives*. Basingstoke: Palgrave Macmillan.
- Talbot, Kyle & Sarah Mercer. 2018. Exploring university ESL/EFL teachers' emotional well-being and emotional regulation in the United States, Japan, and Austria. *Chinese Journal of Applied Linguistics* 41. 410–432. Doi: [10.1515/cjal-2018-0023](https://doi.org/10.1515/cjal-2018-0023)
- Walkinshaw, Ian & Duongthi Hoang Oanh. 2014. Native and non-native English language teachers: Student perceptions in Vietnam and Japan. *SAGE Open* 4. Doi: 10.1177/2158244014534451
- Ziegert, Jonathan C. & Paul J. Hanges. 2005. Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, 90(3), 553–562. Doi:[10.1037/0021-9010.90.3.553](https://doi.org/10.1037/0021-9010.90.3.553)

---

<sup>i</sup> We do not wish to perpetuate the use of the terms NS and NNS but we will use them when discussing studies that used the terms or when they were used in our own experiment. Our preferred terms of L1 and LX will be used in our own discourse whenever they are appropriate. We hope the reasoning for this usage is clear from the article.

<sup>ii</sup> We deliberately used the terms in the study to generate data as it is the bias underlying exactly these terms we wish to examine. However, below we explain how students were debriefed fully to counter the use of the terms for the purposes of the study.